www.bsc.es

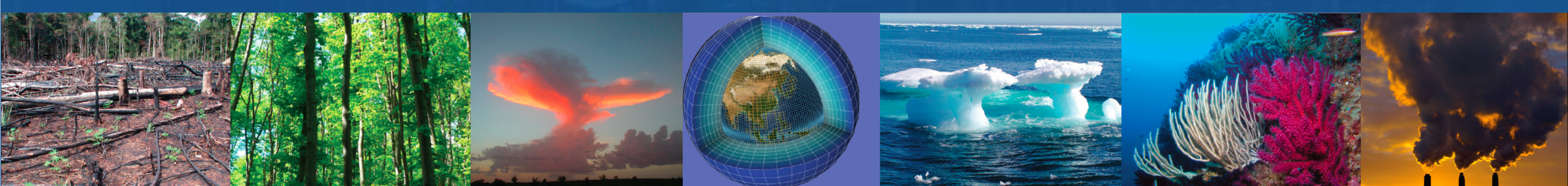
VI JORNADA COMPLEXITAT.CAT
May 2017, Castelldefels

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

# Scientific & Computing challenges in Climate Modeling


Laurent Brodeau, BSC / Earth Sciences dept.

# Context

**Present day climate models have become monsters:**

- extremely complex and heavy to handle
- tremendously greedy in terms of
  - Computing resources (thousands of processors / simulation)
  - Human work force / collaboration

→ **energy & money**

- their results is still highly uncertain

→ Climate modelers warm the planet! For nothing?

# Context

They do warm the planet, but not for nothing!

Climate models are the only tools we have to getting a bit nearer to **understand**:

- the complexity of the Earth's climate system
- how the near future might look like in a GW context

→ ANTICIPATE

→ TRIGGER ACTION (decision makers)

# Outline

## I. Climate models: concept and evolution

- from standalone ocean/atmosphere models to Earth System Models

## II. Challenges of climate modeling

- technological
- scientific

## What is a GCM (global General Circulation Model) ?

A set of mathematical equations that represent the **flow** and **thermodynamics** of the whole 3D **Atmosphere**, or **Ocean** (AGCM and OGCM, respectively).

Today's GCMs, are typically run at a horizontal resolutions from 100km down to a 1 km. They can only be run on **supercomputers** → using typically 200 up to 10 000 processors at the same time.

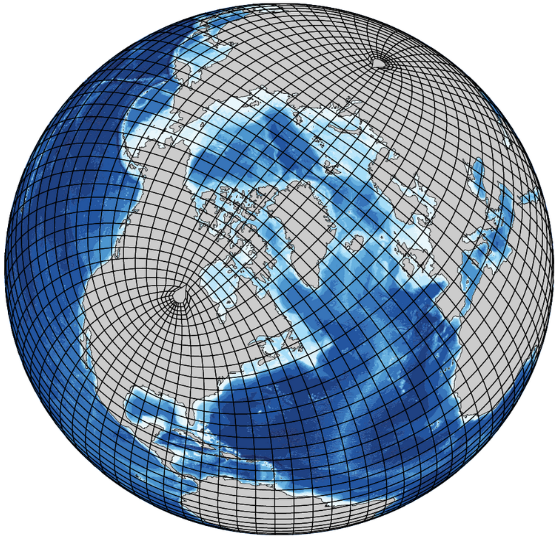
# Global Circulation Models (GCMs)

## Concept, over-simplified

**1.** Equations that describe your process of interest (laws of physics):

$$\frac{\partial T}{\partial T} - \alpha \frac{\partial^2 T}{\partial x^2} = 0$$

**2.** Build your gridded domain:



**3.** Discretize your equations accordingly:

$$T_i^{n+1} = T_i^n + \alpha \frac{\Delta t}{\Delta x^2} \left[ T_{i-1}^n - 2T_i^n + T_{i+1}^n \right]$$

**4.** Solve in space (3D) and integrate in time for each grid point of your gridded domain.

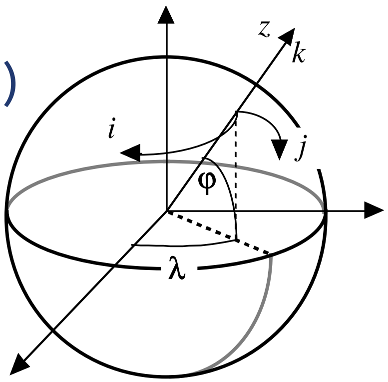
An actual GCM has a number of prognostic variables that are directly integrated (e.g. , pressure; winds/currents; temperature, humidity/salinity) together with a number of diagnostic variables which are deduced from these.

# Ocean GCMs > Governing equations

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

$$\frac{\partial \mathbf{U}_h}{\partial t} = -\left[ (\nabla \times \mathbf{U}) \times \mathbf{U} + \frac{1}{2} \nabla \left( \mathbf{U}^2 \right) \right]_h - f\, \mathbf{k} \times \mathbf{U}_h - \frac{1}{\rho_o} \nabla_h p + \mathbf{D}^{\mathbf{U}} + \mathbf{F}^{\mathbf{U}}$$

(1)




$$\frac{\partial p}{\partial z} = -\rho\, g \qquad (5)$$

$$\nabla \cdot \mathbf{U} = 0 \qquad (6)$$

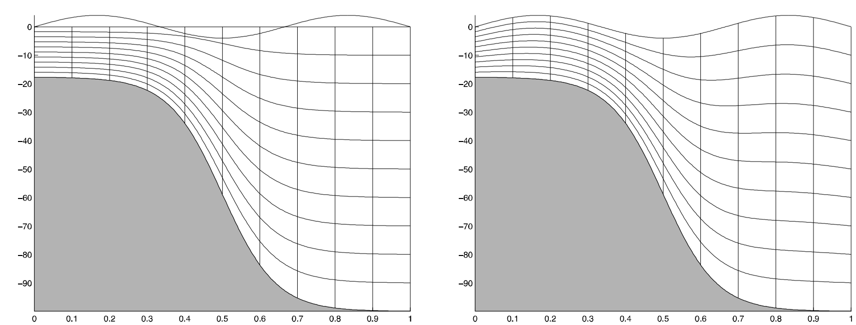
$$\frac{\partial T}{\partial t} = -\nabla \cdot (T\, \mathbf{U}) + D^T + F^T$$ (conservation of heat)

$$\frac{\partial S}{\partial t} = -\nabla \cdot (S\, \mathbf{U}) + D^S + F^S$$ (conservation of salt)

$$\rho = \rho\,(T, S, p)$$ (equation of state)

**Navier-Stockes eq. + hypotheses:**

- (1) Spherical earth approximation
- (2) Thin-shell approximation
- (3) Turbulent closure hypothesis
- (4) *Boussinesq* hypothesis
- (5) Hydrostatic hypothesis
- (6) Incompressibility hypothesis



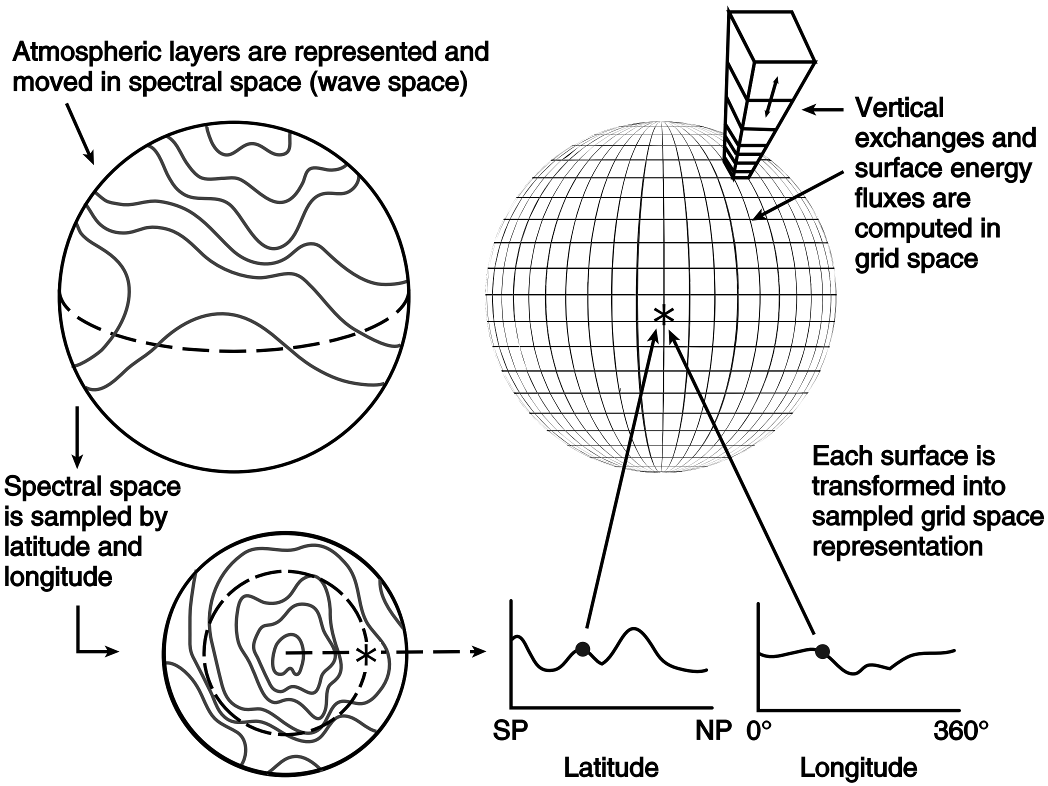
Vertical discretization

# Atmosphere, specific aspects (w.r.t, ocean)

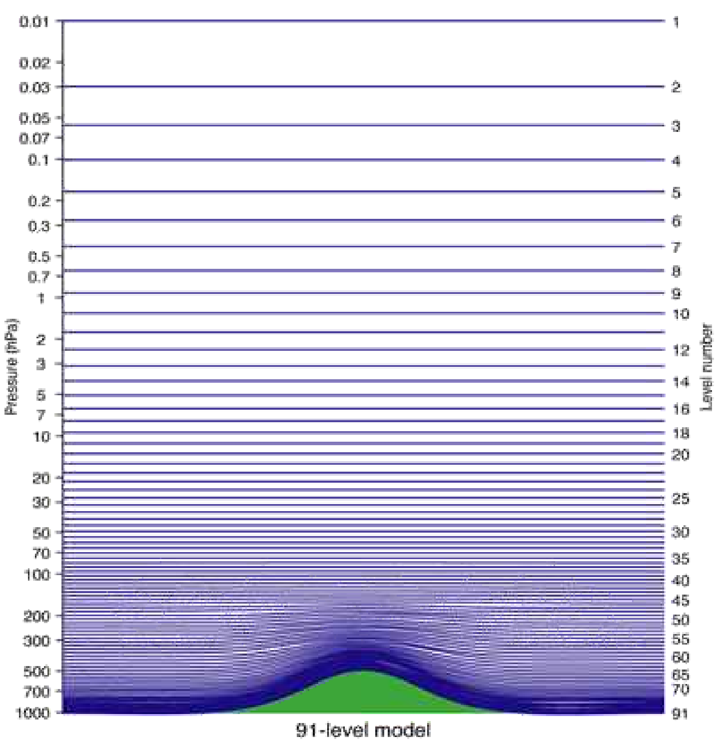


**Horizontal:**
**Spectral** rather than Cartesian

**Time integration:**
**Semi-lagrangian** rather than finite difference (Euler, Leap-frog)




**Vertical:**
**Hybrid coordinates**



## Due to some differences between oceanic and atmospheric flow properties, OGCMs and AGCMs have evolved separate ways

### Ocean

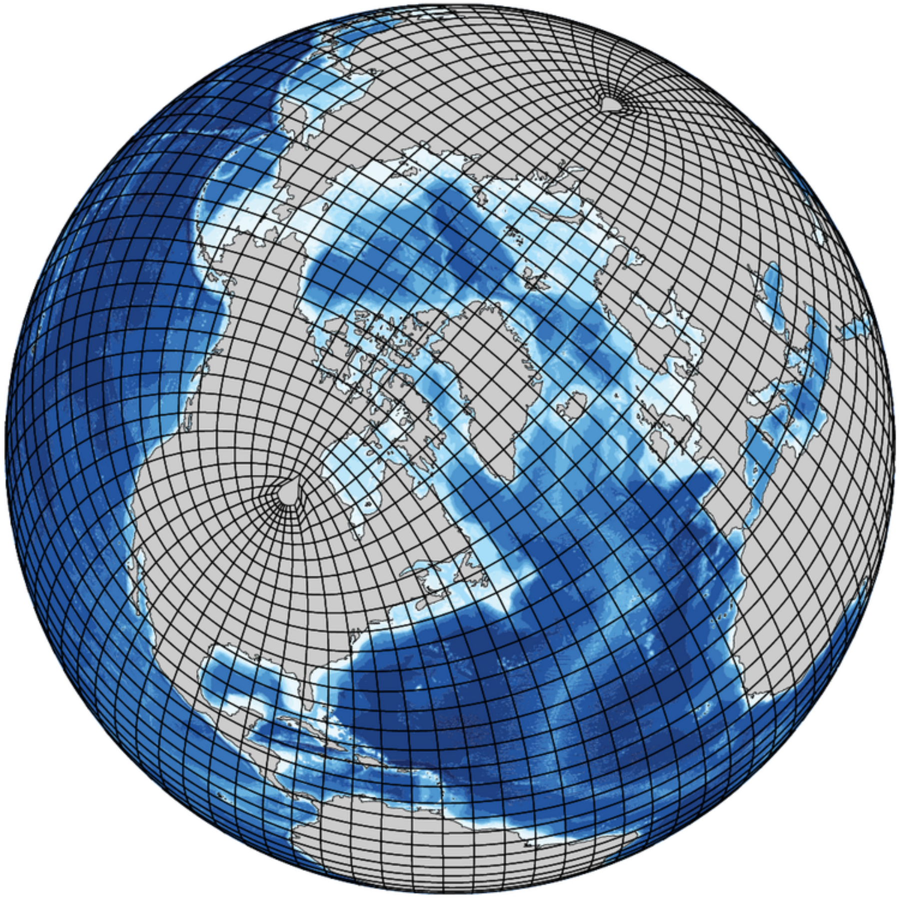
- Much higher resolution needed to solve mesoscale eddies
- Boundary conditions (coastline)
- Must avoid grid singularities at the North pole
- Horizontal: Cartesian, Finite differences,
- Vertical: z, sigma coordinates,
- Time integration: finite difference

### Atmosphere

- Horizontal: spectral representation
- Time integration: semi-Lagrangian

### Common

- Hydrostatic assumption



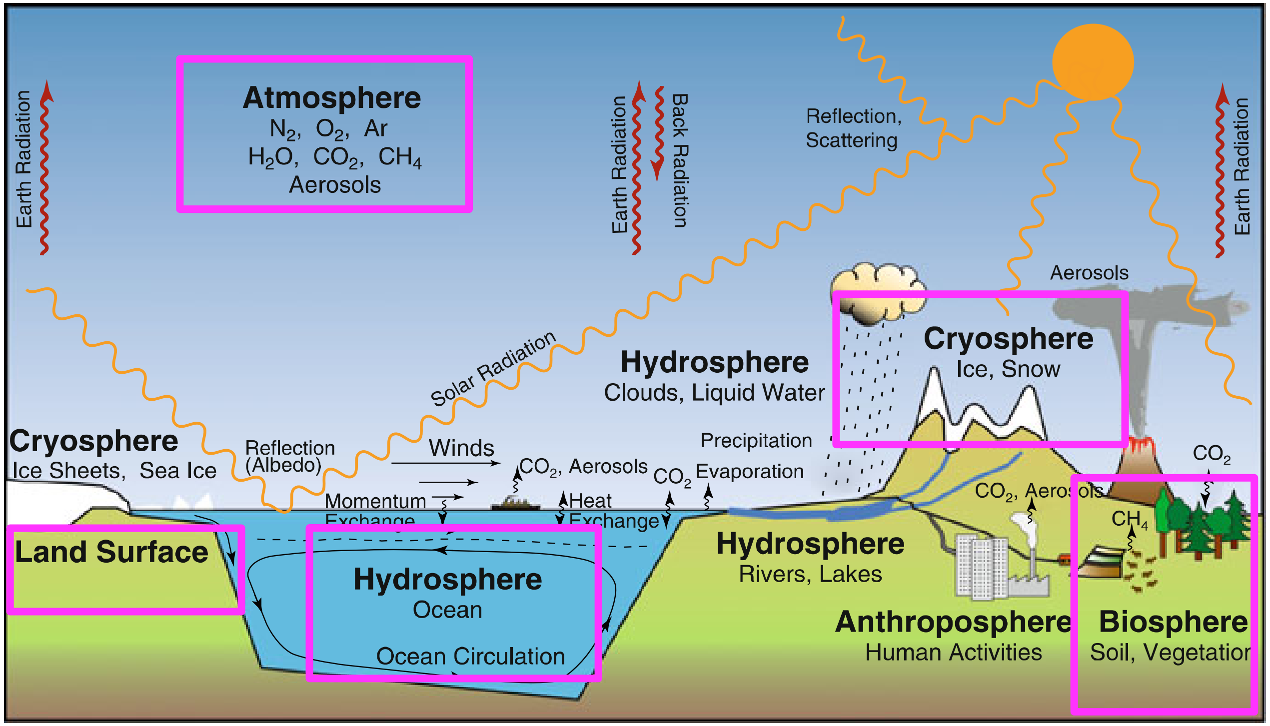
# Climate a.k.a "Earth System" Models (ESMs)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

## What is a climate model ?

Originally, it was an AGCM coupled to an OGCM.

The **climate model** of today is what we now refer to as an "**Earth System Model**" (**ESM**), *e.g.* a collection of **inter-coupled numerical models**, each simulating a given **component of the Earth climate**.



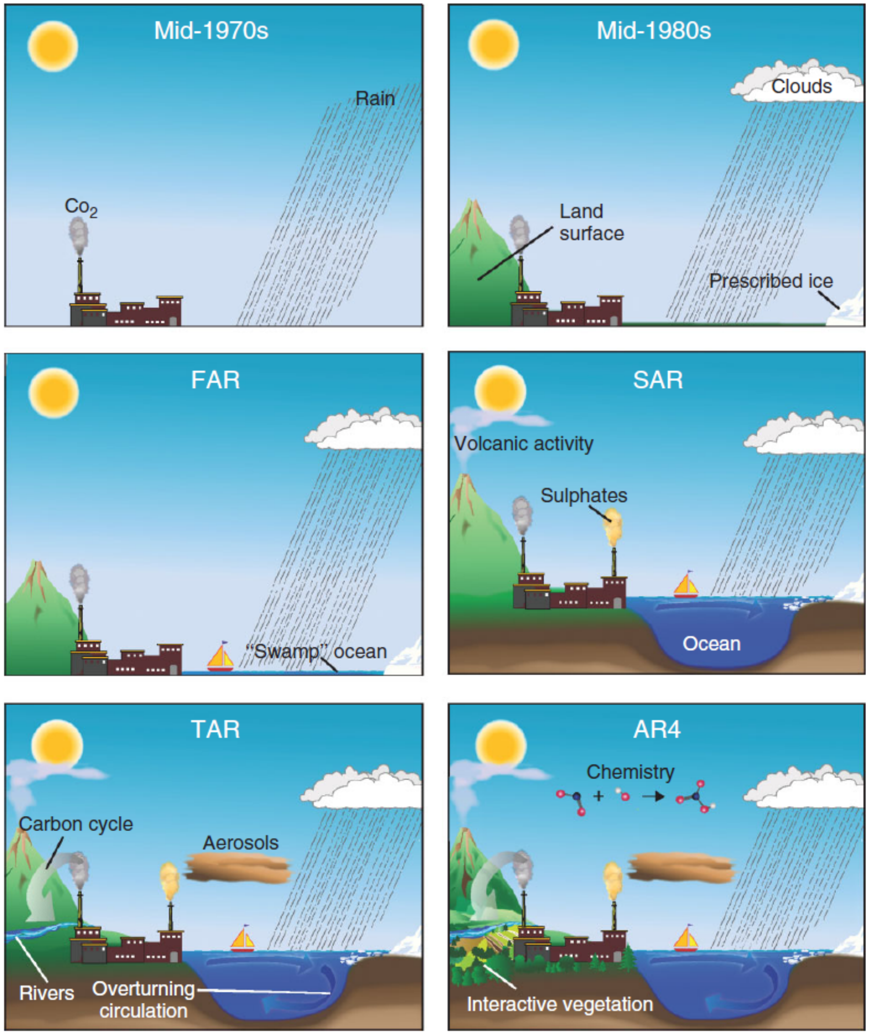
# From standalone AGCMs to Earth System Models (ESMs)

- **It all started with AGCMs (1970s)**
Meteorology, Numerical Weather Prediction

- **Then came the OGCMs (1980s)**
Thermohaline circulation, meridional, overturning circulation, etc.

- **Then coupled AGCM+OGCM=AOGCMs (1990s)**
sea-ice model in OGCMs
The first climate models

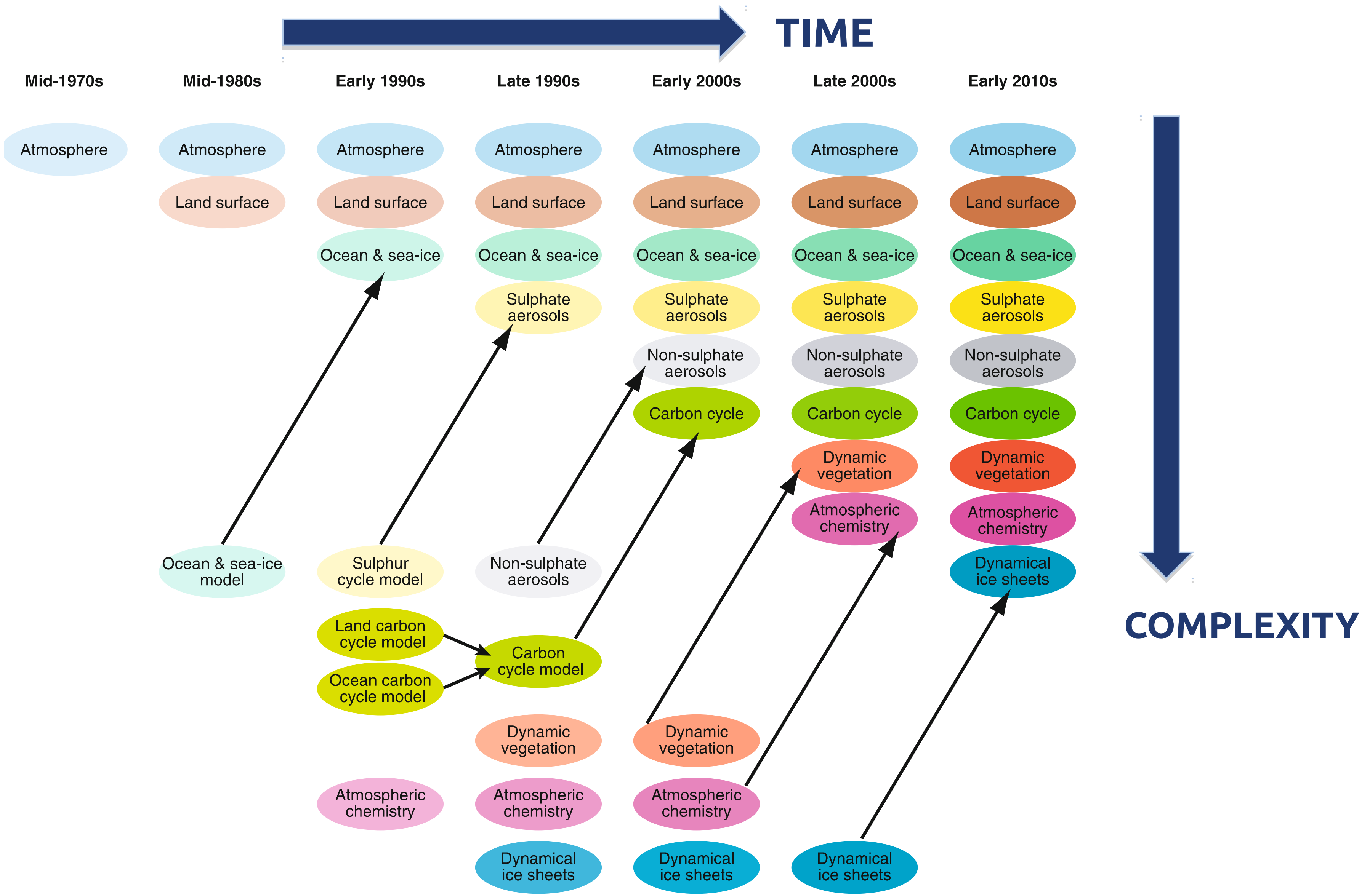
- **Towards true ESMs (2000s)**
Inclusion of remaining climate components (land, atmospheric chemistry, vegetation, ice sheets, carbon cycle



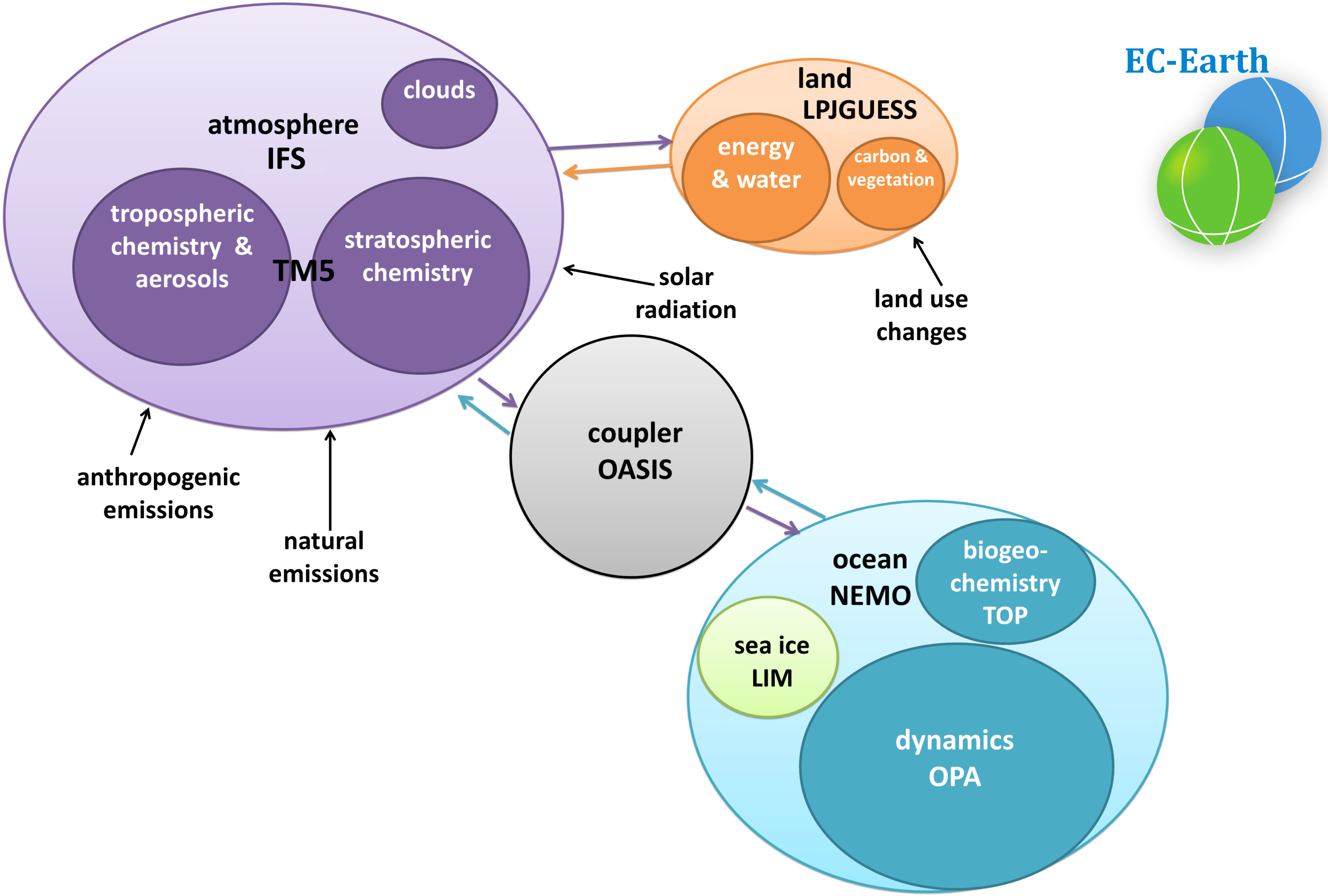
# From standalone GCMs to Earth System M.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA



# EC-Earth: typical ESM of today



# II. Challenges in Climate modeling

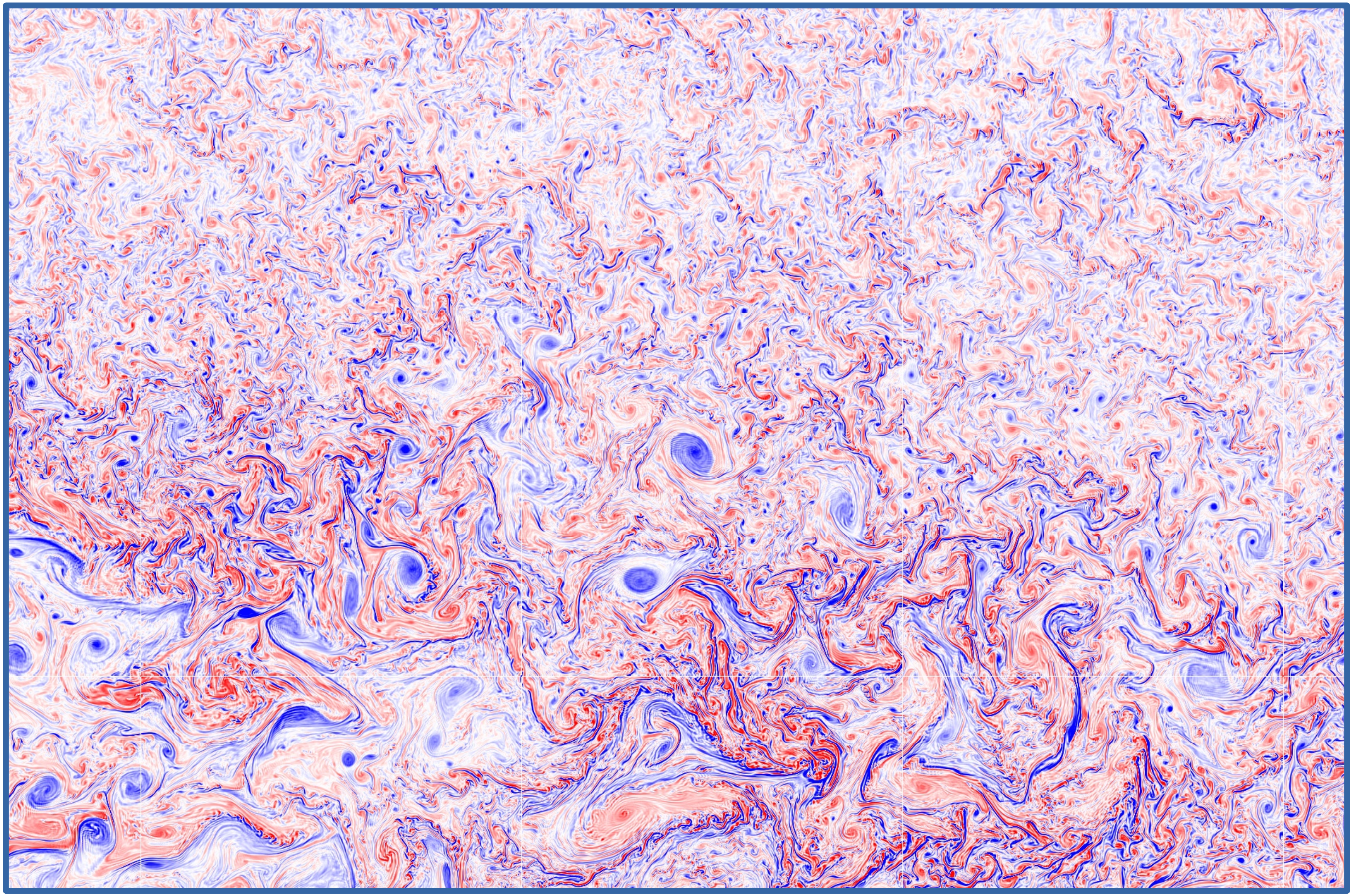
- **Technological challenges:**

  Why: Increase of resolution, ensemble size & length of the simulations

  - HPC technologies ! (exaflop computing, data storage, CPU/GPU, etc.)
  - Improve scalability (performance) of the models

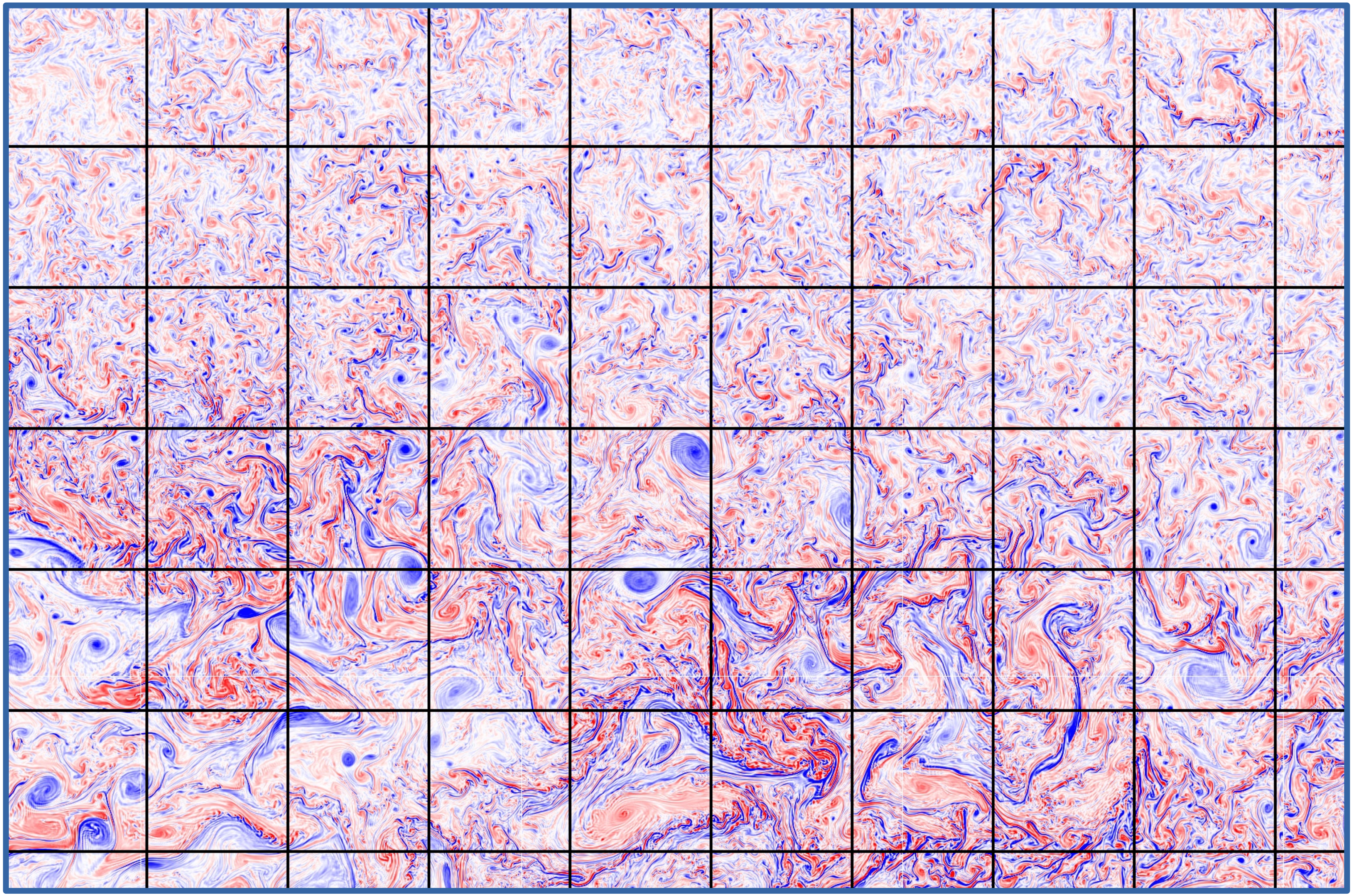
- **Scientific challenges**
  - Better representation of subgrid-scale processes
  - Prediction uncertainty formulation via probabilistic approach (ensemble of simulations)

# Model resolution and subgrid features



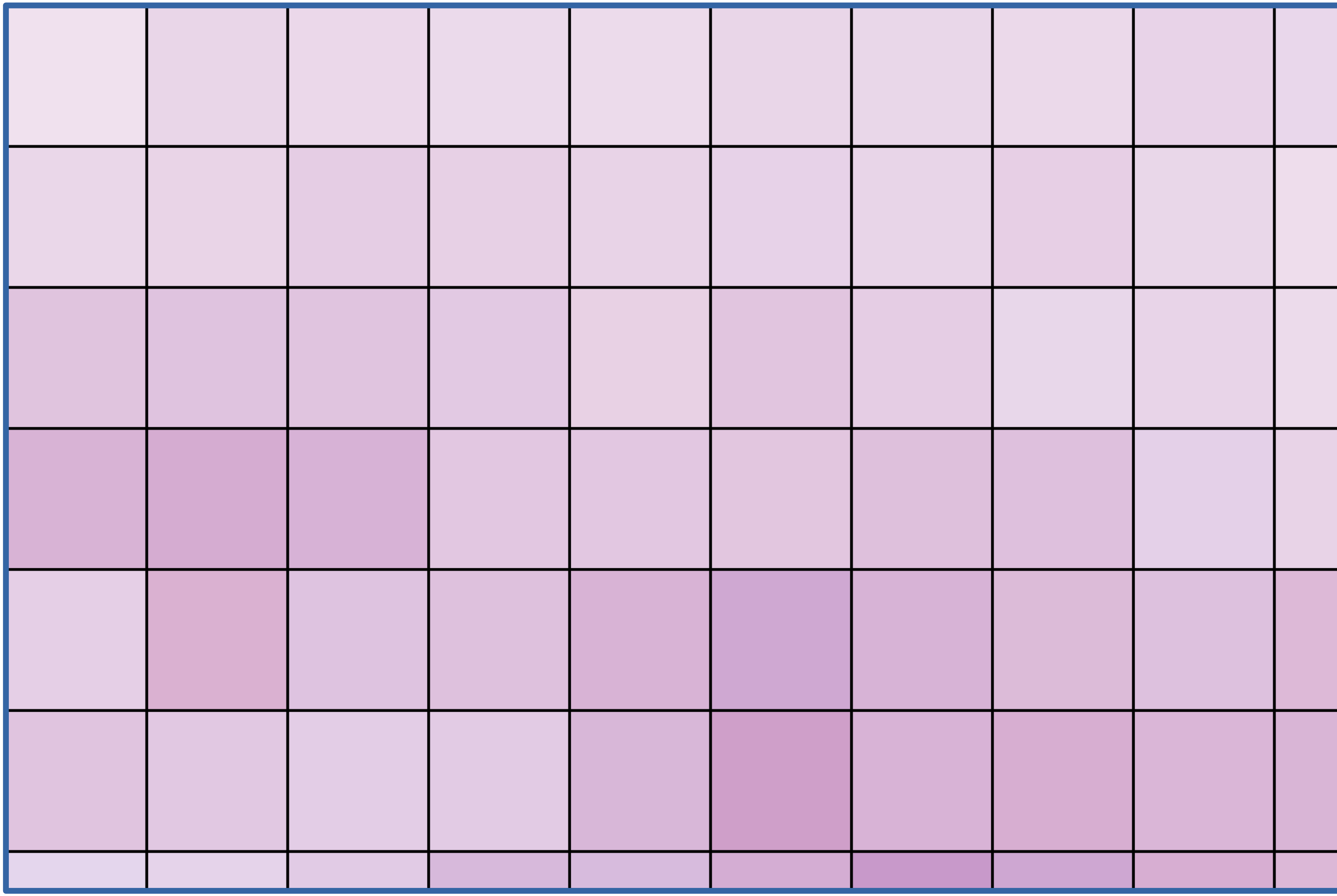
# Model resolution and subgrid features

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

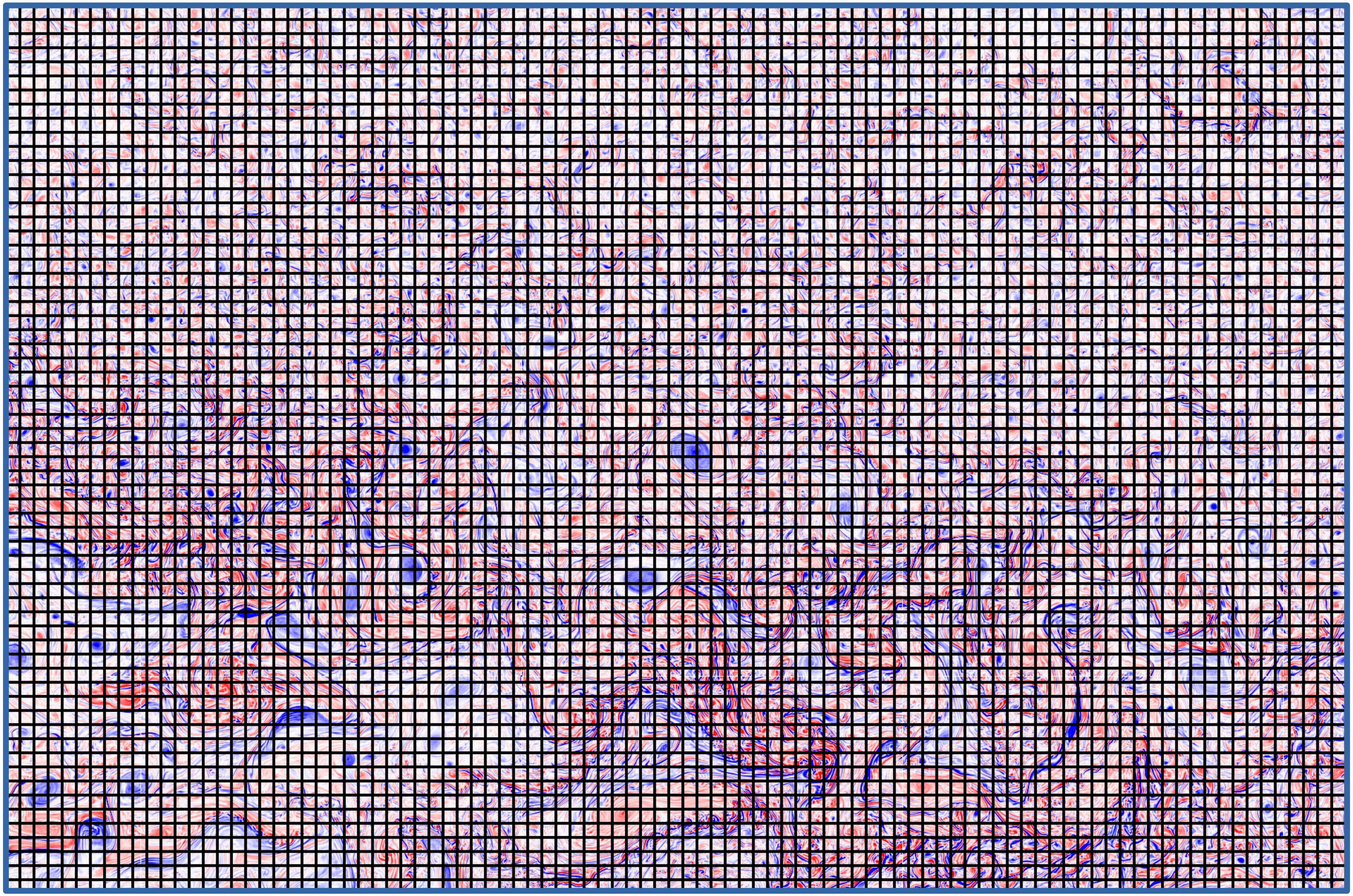


# Model resolution and subgrid features

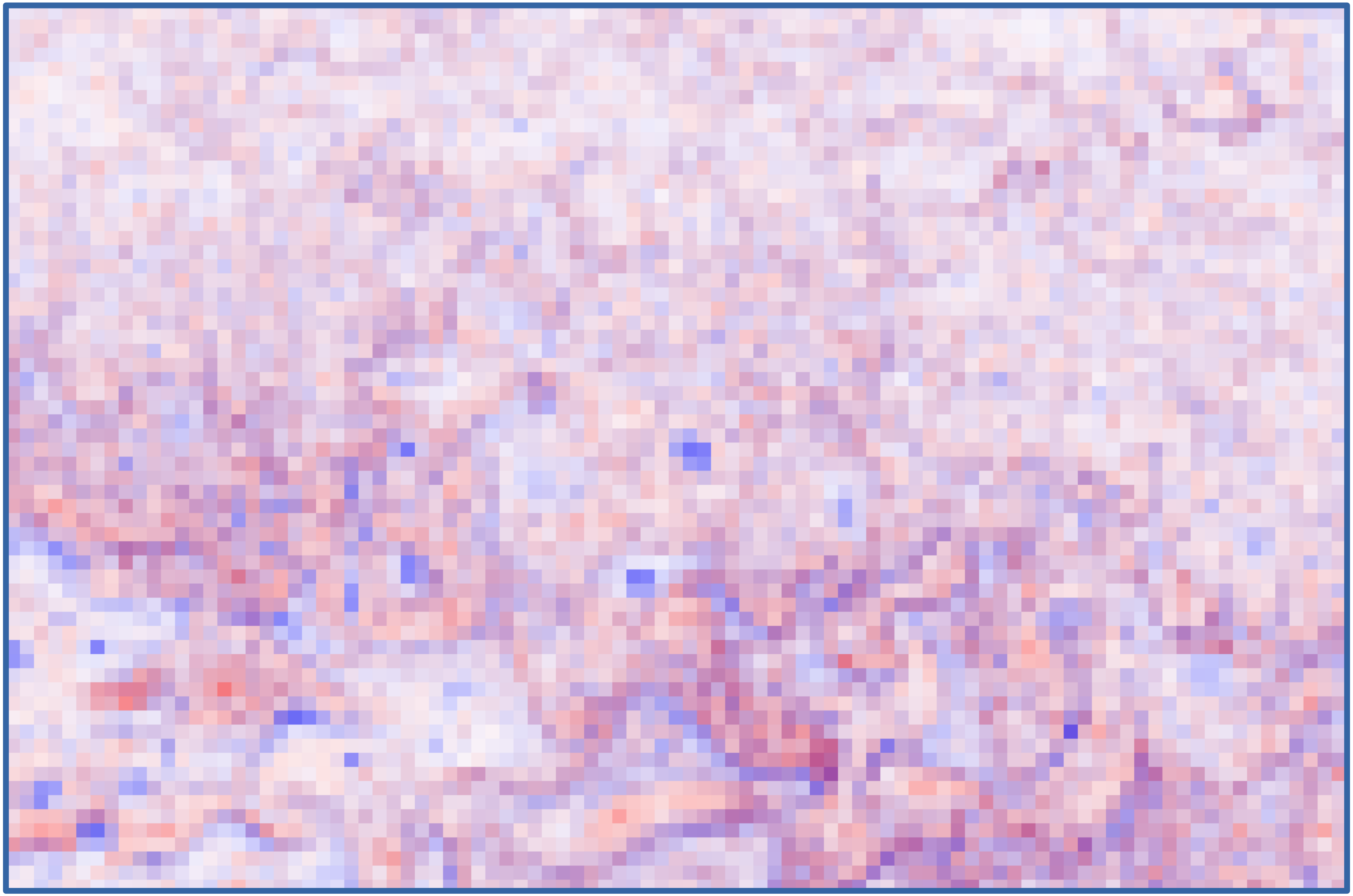




# Model resolution and subgrid features



# Model resolution and subgrid features



# Our resolution-based problem:

The resolution of the gridded computational domains we use will hardly be thin enough to resolve the processes of interest (like sub-mesoscale eddies in the ocean, convection in atmosphere).

## Solution 1:

- Keep on increasing the spatial resolution of models

  → resolve more features, rely less on parametrization, resolve convection (dream on: <=1km)

  → exponential needs in HPC resources

  **Problem #1**: we don't have the HPC resources available yet!

  → exaflop supercomputers! (currently petaflop)

  **Problem #2:** even if we had, our ESMs wouldn't take advantage of them because of the (awful) way they were assembled during the past 3 decades! (by geo-scientists, not computer scientists)

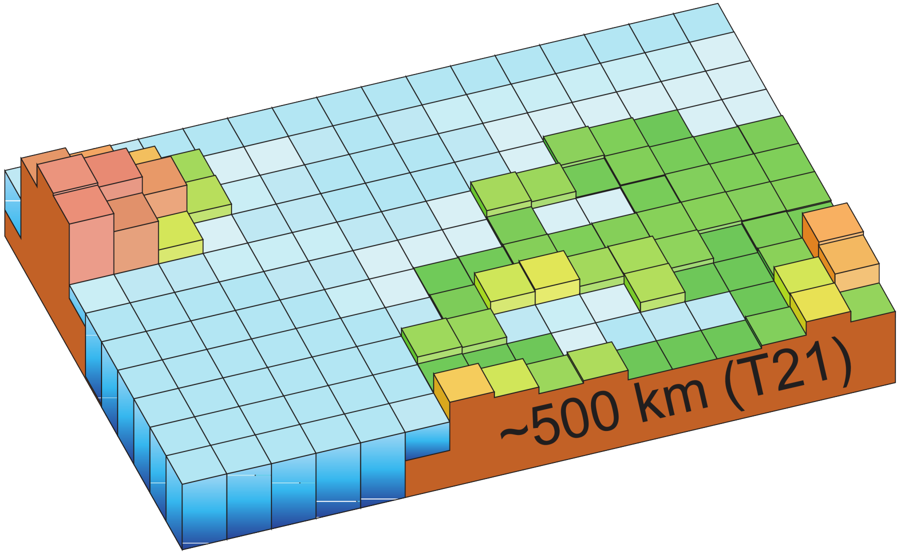
# Resolution and subgrid-scale processes



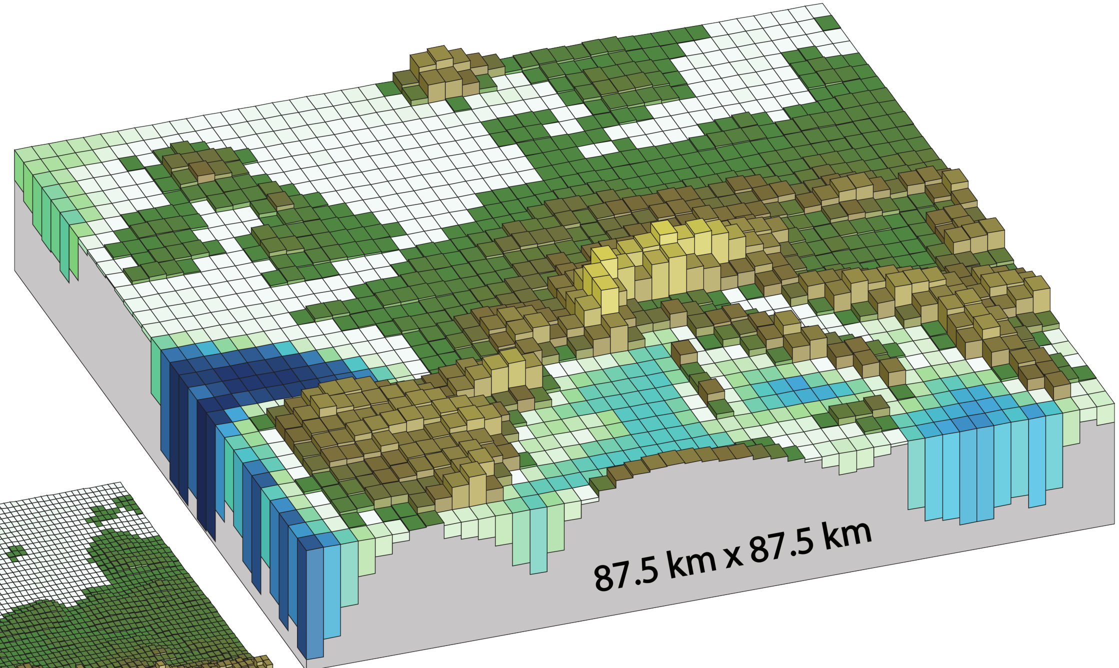
## Solution 2: 

- Improve the parametrization of subgrid-scale processes

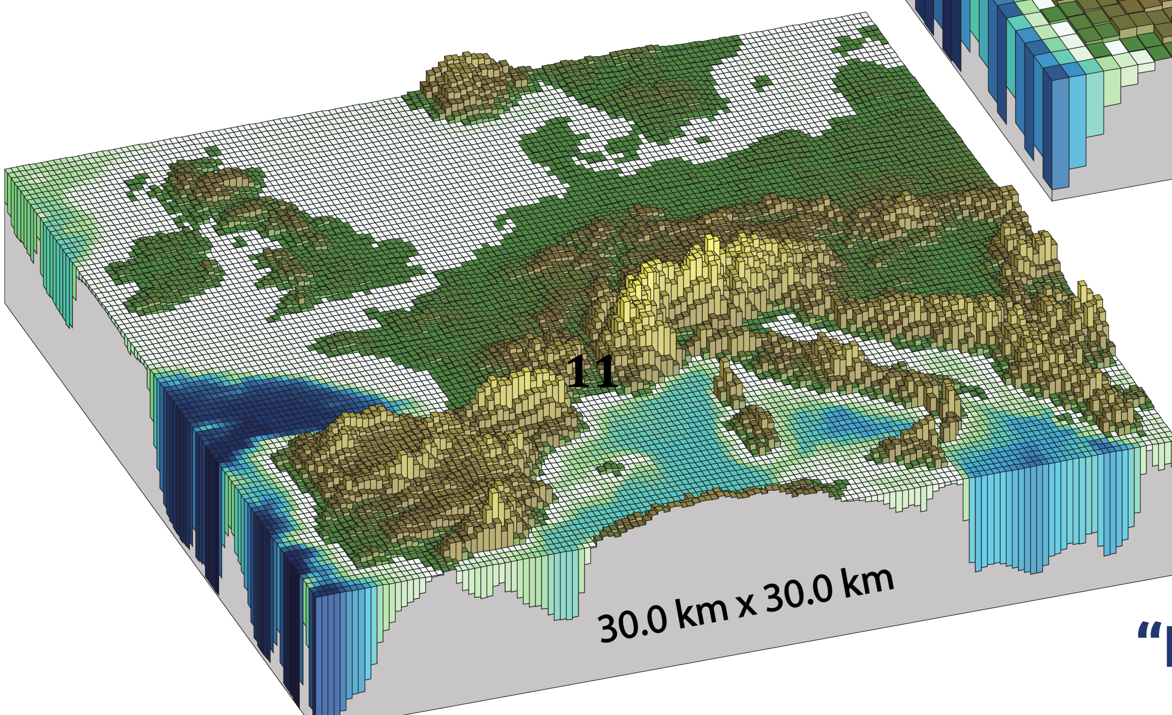
# Zoom over Europe in a typical climate model of the 1990's (500km resolution)



# Today, CMIP6 (ongoing)

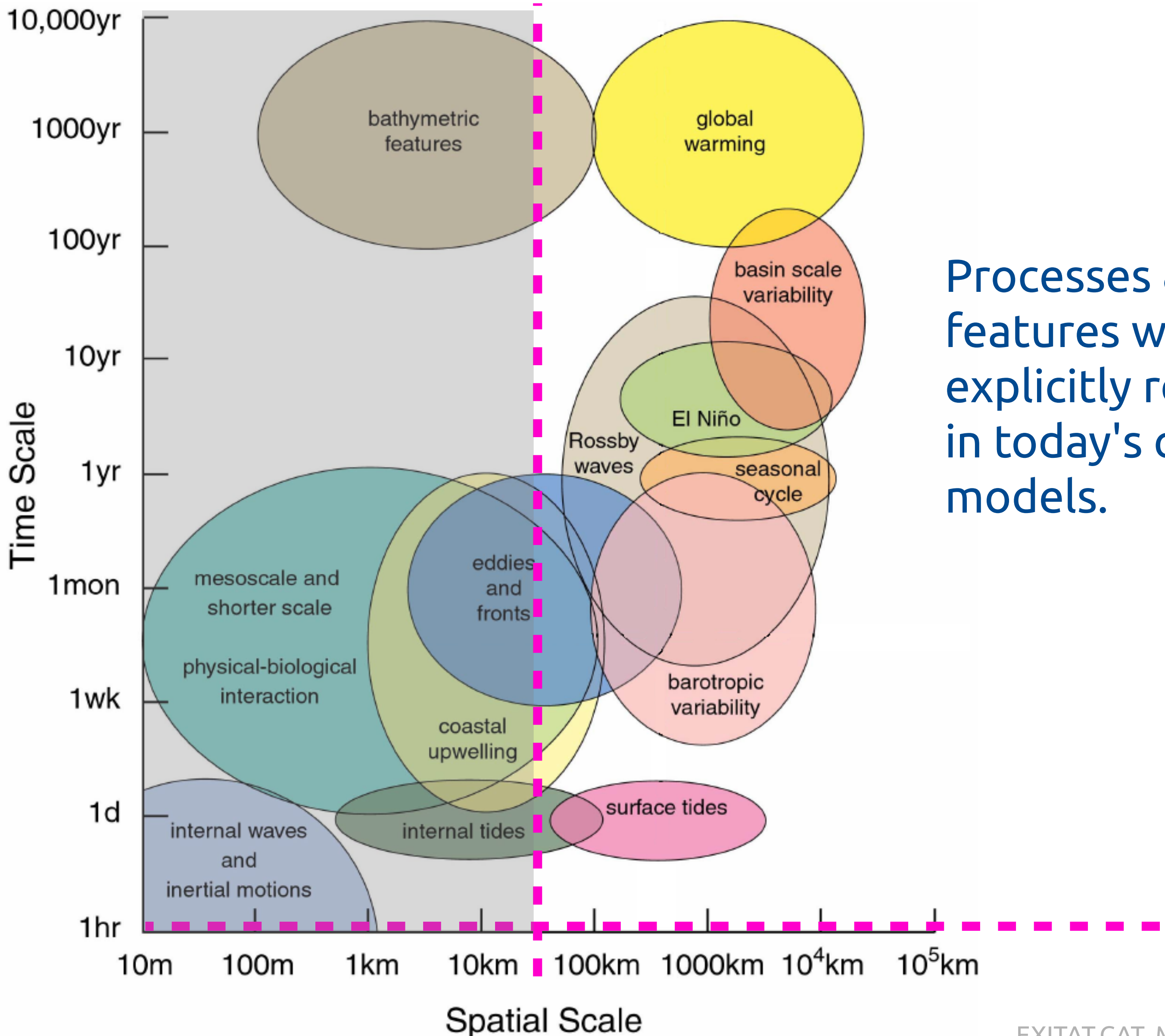



**Default resolution**




**“High” resolution**

# Resolution: Climate models of today



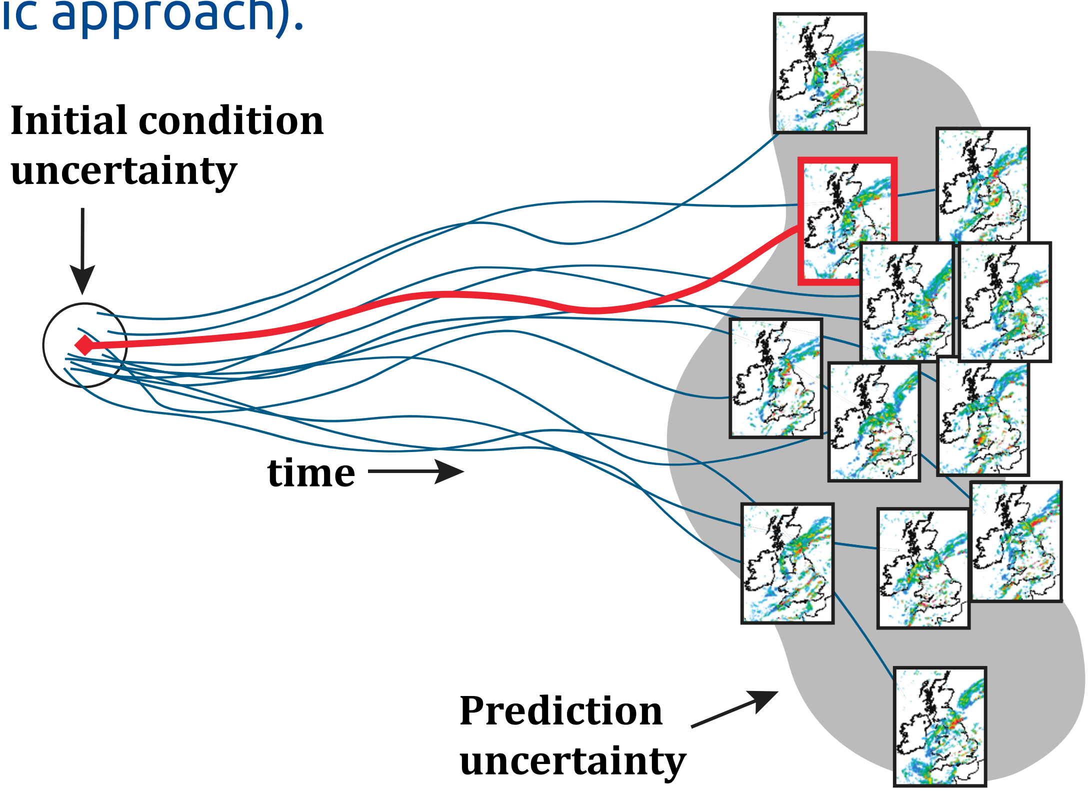

Processes & features we can explicitly resolve in today's climate models.

# Ensemble simulations

## Probabilistic rather than deterministic

Concept: draw your conclusions based on statistics carried out on an ensemble of N "almost-identical" simulations rather than 1 unique simulation (deterministic approach).

**N**: the bigger the better, typically N=50!




*Bauer et al, 2015, Nature*

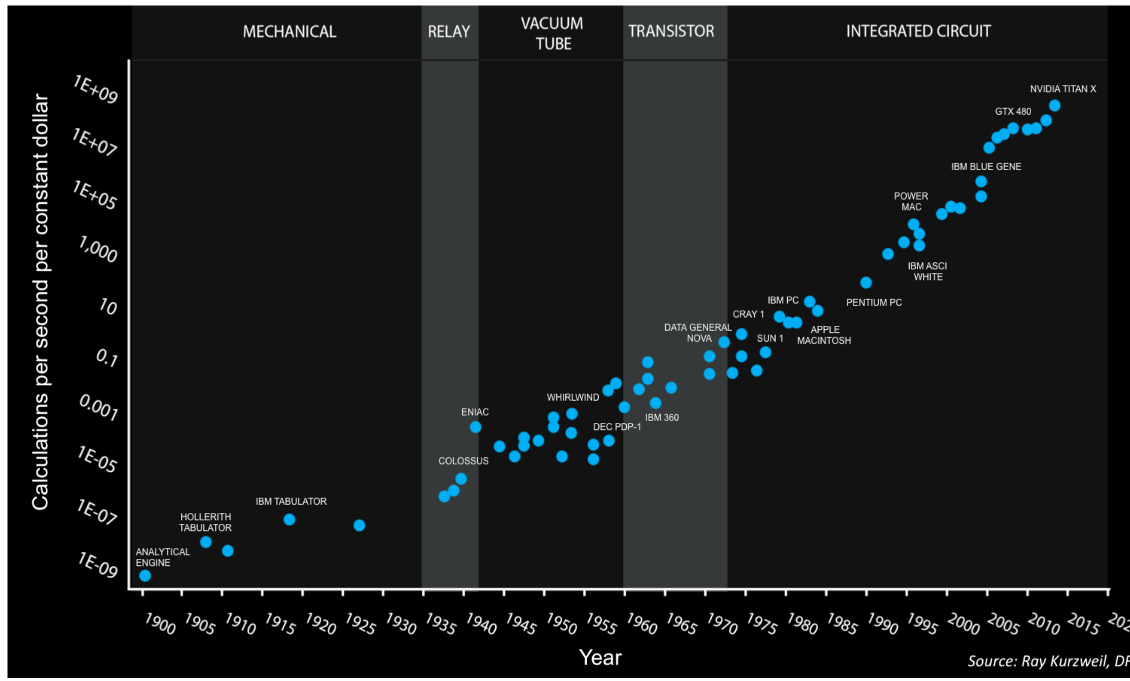
# Model needs vs compute growth

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Compute growth (Moore's Law) is keeping up with increasing ESMs requirements

## Limitations are:

- weak scalability* capabilities of ESMs
- data storage growth is not keeping up




* Scalability: computing efficiency of the model as the number of processors used (in parallel) increases

# New supercomputing technologies

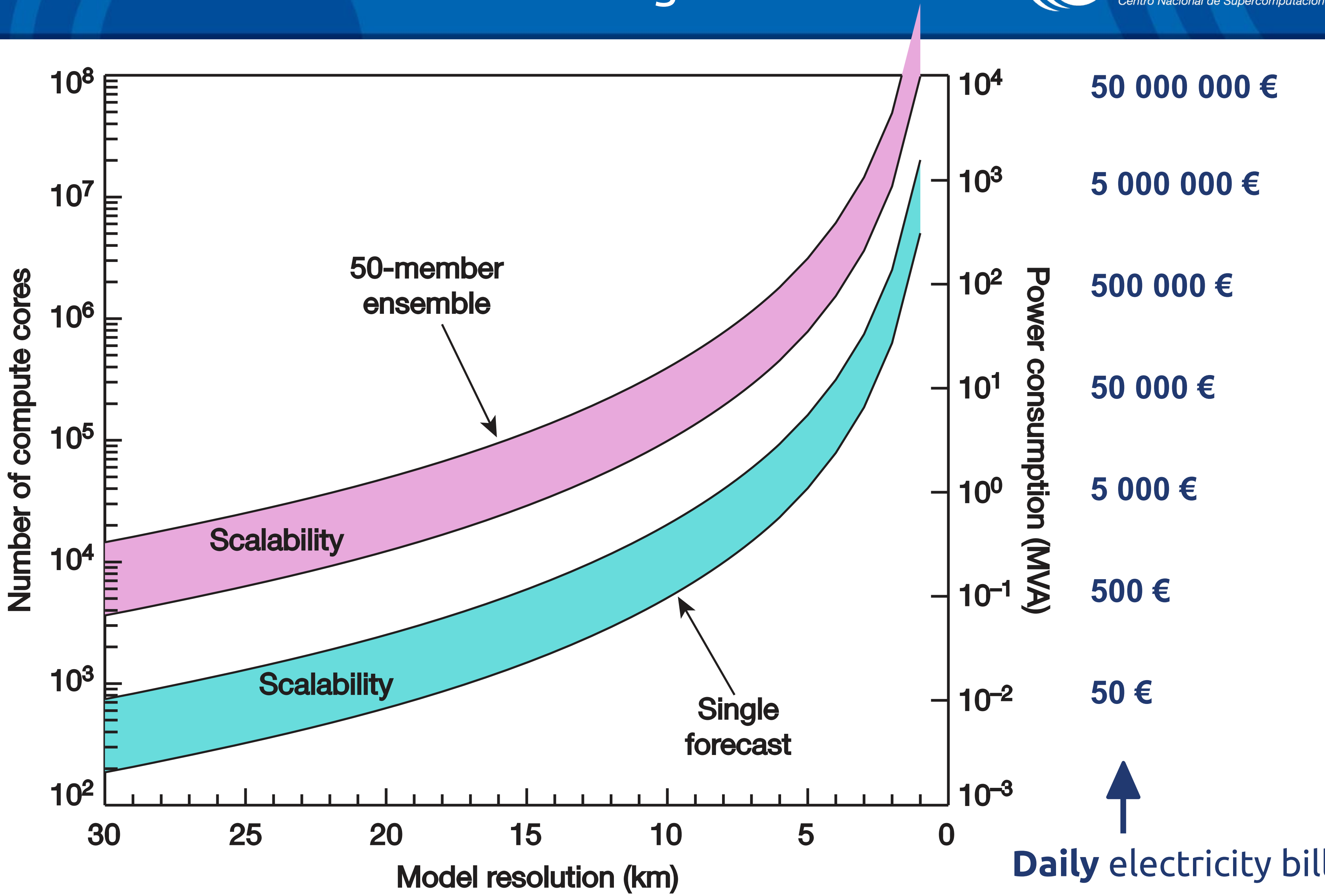
**Hardware** is not initially designed for climate models
Supercomputers are rarely built for climate models…
→ need for supercomputers more adapted to ESM needs

**Ongoing work:**

- Supercomputers that combine both CPUs and GPUs, to take advantage of the best of both worlds depending on the given ESM routine/module in use

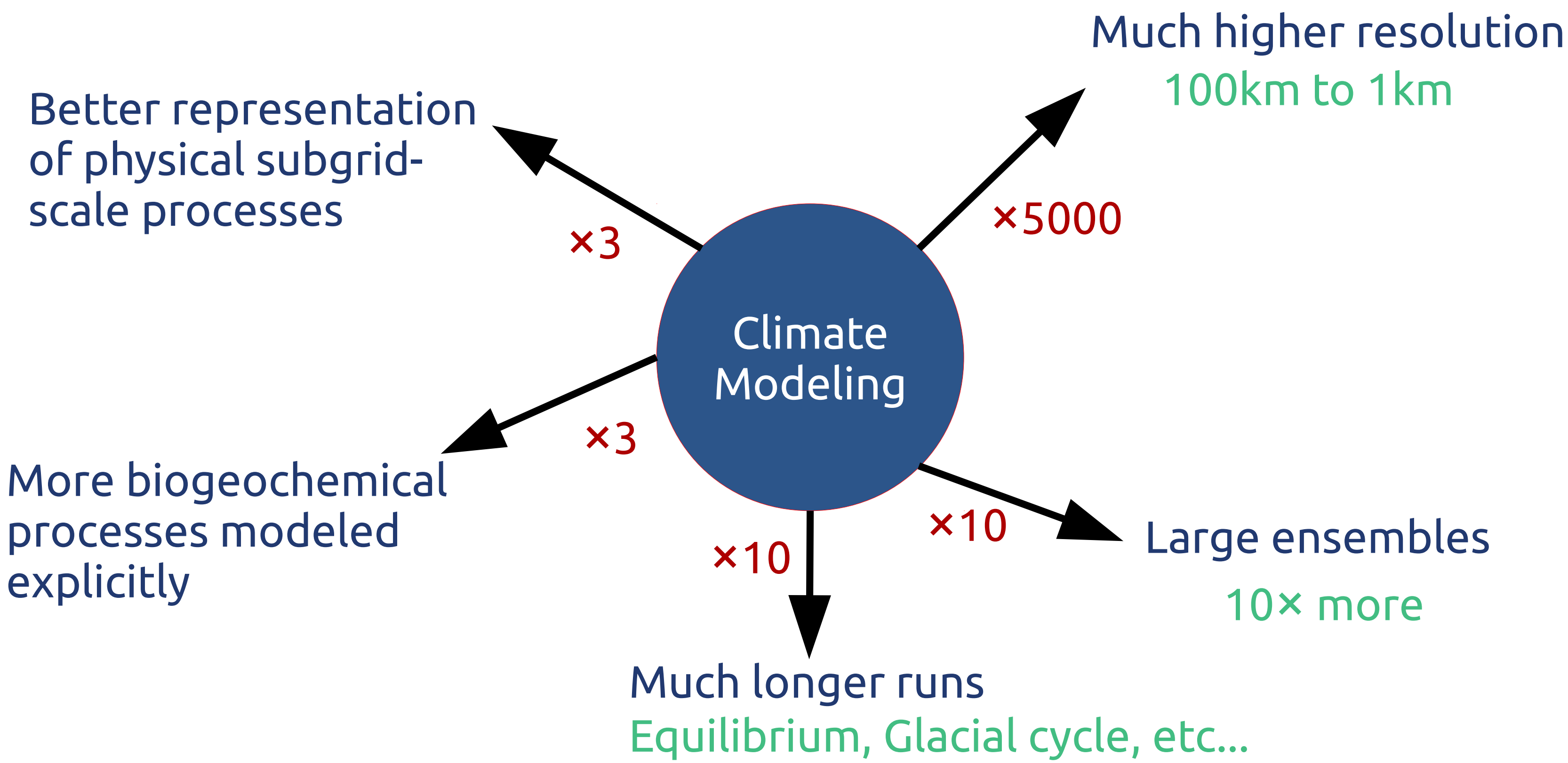
Why not waiting a bit longer for simulations to finish if they cost much less?
→ use cheaper and more economic processors like mobile chips (ARM)

# The cost of better simulating Earth




*Bauer et al, 2015, Nature*

# The cost of better simulating Earth

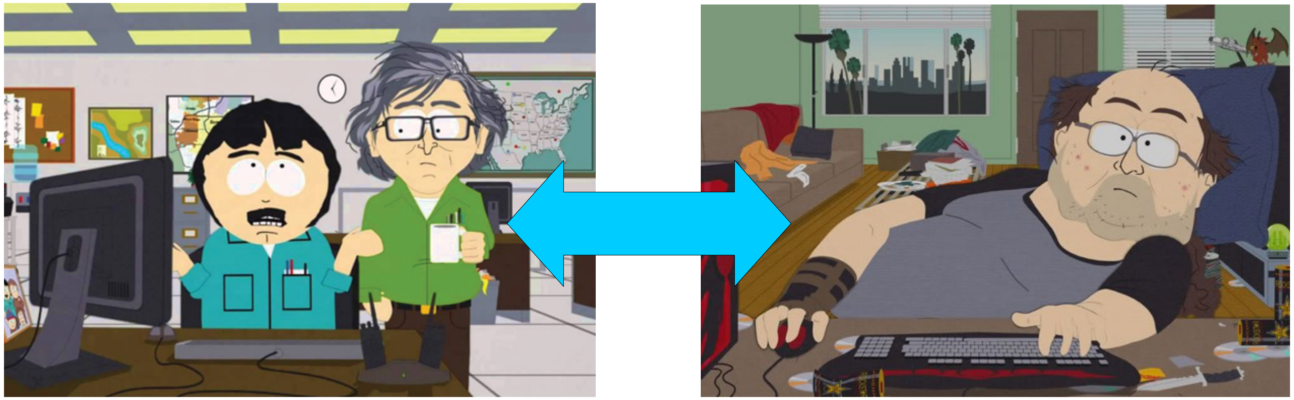



**→ Aggregate cost ~ $10^6$**

*"Computing & data handling will be the #1 challenges for weather and climate prediction in the next 10 years"*

## We need:

- Re-write ESMs from scratch so that they decently scale, so we can keep on increasing the resolution of the models
  → computer scientists wanted !
- More powerful supercomputers, adapted to ESM needs, also hopefully less greedy in electricity!
- **Way more** data storage capacity



Climate scientists    Computer scientists

www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

# Thank you!